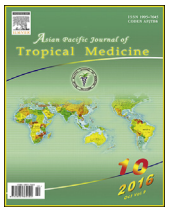
HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

Asian Pacific Journal of Tropical Medicine

journal homepage: http://ees.elsevier.com/apjtm



Original research http://dx.doi.org/10.1016/j.apjtm.2016.07.024

Identification and analysis of a processed cytochrome P450 pseudogene of the disease vector *Aedes aegypti*

Fatma M.A. El-garj, Mustafa F.F. Wajidi, Silas W. Avicor✉

Molecular Entomology Research Group, School of Distance Education, Universiti Sains Malaysia, 11800 Minden, Penang, Malaysia

ARTICLE INFO

Article history:
Received 13 May 2016
Received in revised form 10 Jun 2016
Accepted 15 Jul 2016
Available online 25 Aug 2016

Keywords:
Aedes aegypti
Clone
Cytochrome P450 pseudogene
CYP4H44P
Bioinformatics

ABSTRACT

Objective: To clone cytochrome P450 from *Aedes aegypti* (*Ae. aegypti*) and determine the characteristics using bioinformatics tools.

Methods: Cytochrome P450 of *Ae. aegypti* was amplified using polymerase chain reaction, cloned and sequenced. Evolutionary relationship of the sequence was inferred and bioinformatics tools were used to predict subcellular localisation, signal peptide, transmembrane helix, phosphorylation, O-glycosylation, secondary and tertiary structures of the deduced protein.

Results: Polymerase chain reaction rather amplified a cytochrome P450 pseudogene which was named *CYP4H44P* (GenBank accession number KF779932). The pseudogene has 1 537 nucleotides and an open reading frame of 335 amino acids containing cytochrome P450 motifs except the WxxxR motif. It is highly homologous to *CYP4H28* and *CYP4H28v2*. Phylogenetic analysis and evolutionary divergence showed strong clustering with *CYP4H28* alleles and least divergence from the alleles respectively. The deduced protein was predicted to be found in the cytoplasm and likely to be phosphorylated but devoid of signal peptide, transmembrane helix and O-glycosylated sites. The secondary and tertiary structures were also generated.

Conclusions: A cytochrome P450 pseudogene, *CYP4H44P* was cloned from *Ae. aegypti*. The pseudogene is homologous with *CYP4H28* alleles and seems to have recently diverged from this group. Isolating this pseudogene is an important step for evaluating its biological role in the mosquito and for the evolutionary analysis of *Ae. aegypti* CYPs.

1. Introduction

Cytochrome P450 monooxygenases (CYPs) are an important enzymatic super group performing diverse functions in different organisms [1,2]. The group is made up of different clans which are subdivided into families [1]. CYPs perform a broad range of functions in insects, including insecticide resistance and activities related with insect physiology [1]. CYP genes and pseudogenes have been identified in insect genomes. Pseudogenes constitute about 3.1% of the CYPs in the genome of the yellow fever mosquito, *Aedes aegypti* (*Ae. aegypti*) [3]. Although similar to functional genes, pseudogenes are regarded to have lost the ability to code for functional proteins [4,5]. They are grouped into three types, namely processed pseudogenes, duplicated pseudogenes and unitary pseudogenes [6,7]. Processed pseudogenes are formed as a result of retrotransposition of mRNA into the genome [6,7]. The duplicated pseudogenes are formed when functional genes duplicate, and one of the duplicates undergoes mutation and becomes non-functional while unitary pseudogenes arise when there is a disruptive mutation in the coding genes of functional proteins [4,6,7]. Despite being presumed as not coding for functional proteins, pseudogenes can produce transcriptional products and perform several roles in organisms [6–8]. During the

First author: Fatma M. A. El-garj, Molecular Entomology Research Group, School of Distance Education, Universiti Sains Malaysia, 11800 Minden, Penang, Malaysia.
Tel: +60 174074032
E-mail: fatmagorj@yahoo.com
✉Corresponding author: Silas W. Avicor, Molecular Entomology Research Group, School of Distance Education, Universiti Sains Malaysia, 11800 Minden, Penang, Malaysia.
Tel: +60 142401941
E-mails: wintuma@live.com, swavicor@usm.my

Peer review under responsibility of Hainan Medical University.
Foundation Project: This work was funded by the Malaysian Ministry of Education (ERGS 203/PJJAUH/6730097) and Universiti Sains Malaysia (RU grant 1001/PJJAUH/815095).

evolutionary process, pseudogenes are presumed to be under less conservation constraints compared to functional genes and hence are useful in analysing the evolutionary history of genomes and functional genes [6]. Interest in pseudogenes has grown and the possibility of artificially synthesising functional translational products from them has even been hypothesised [9].

Ae. aegypti is a vector of arthropod borne viruses such as chikungunya, dengue, yellow fever and Zika. Routine control of the vector using insecticides has contributed to CYP-mediated insecticide resistance. Consequently, CYPs of this mosquito have been studied for their functional roles in insecticide resistance [3,10]. In the course of isolating CYP gene fragments from *Ae. aegypti* [11], a processed CYP pseudogene was identified. This study describes isolation of the pseudogene and the predicted characteristics of its deduced protein. Identification of the CYP pseudogene will be useful in evolutionary analysis of functional CYPs of *Ae. aegypti* and in future studies to determine its biological role in the mosquito.

2. Material and methods

2.1. RNA extraction

Total RNA was extracted from fourth instar larvae (0.25 g) of a reference *Ae. aegypti* strain [12] as described in [11]. The RNA was qualitatively and quantitatively analysed by electrophoresis in a 1% agarose-formaldehyde gel and using a Nanodrop spectrophotometer respectively.

2.2. cDNA synthesis and amplification reaction

Synthesis of cDNA from RNA (5 μg) was performed using RevertAid™ Premium First Strand cDNA Synthesis Kit (Fermentas®) as per the manufacturer's instructions. The cDNA was used in a polymerase chain reaction (PCR) with a pair of primers (Table 1) as previously described [11] in a PTC-100™ Programmable Thermal Controller (MJ Research).

2.3. 3′- and 5′-rapid amplification of cDNA ends (RACE)

2.3.1. 3′-RACE

The primers 3R1 (0.5 μL of 10 μM) and Oligo(dT)$_{25}$V (0.5 μL of 10 μM) were used in a PCR containing cDNA (2 μL), OneTaq Hot Start 2X Master Mix with Standard Buffer (12.5 μL) and sterile distilled water (9.5 μL) at 94 °C for 5 min, 7 cycles of (94 °C/30 s, 43 °C/30 s and 72 °C/1 min), 27 cycles of (94.0 °C/30 s, 51.5 °C/30 s and 72.0 °C/1 min), and finally at 72 °C/10 min. The product was used as template (2 μL) in a nested PCR using 3R1N and Oligo(dT)$_{25}$V as primers with similar reaction volume and thermal condition as stated above.

2.3.2. 5′-RACE

The SMARTer™ RACE cDNA Amplification Kit (Clontech®) was used to obtain the 5′-RACE Ready cDNA according to the manufacturer's instructions. 5′-RACE PCR was performed using Advantage® GC 2 PCR Kit (Clontech®) according to the manufacturer's instructions at 5 cycles of 94 °C/30 s and 72 °C/3 min, followed by 5 cycles of (94 °C/30 s, 68 °C/30 s and 72 °C/3 min) and then 30 cycles of (94 °C/30 s, 66 °C/30 s and 72 °C/3 min).

2.4. Electrophoresis and purification

The amplified product was electrophoresed in a 1% agarose gel and viewed under ultra-violet light. The product was purified using the Wizard® SV Gel and PCR Clean-Up System (Promega®) according to the manufacturer's instructions.

2.5. Cloning and plasmid extraction

The purified product was ligated into a pGEM®-T Easy Vector (Promega®) in an insert: vector ratio of 3:1. The ligation reaction consisting of 3 μL purified fragment, 1 μL of pGEM®-T Easy Vector, 5 μL of 2X rapid ligation buffer and 1 μL of T4 DNA Ligase (3 Weiss units/μL) was incubated overnight at 4 °C. The ligated product was cloned into competent *Escherichia coli* JM 109 after heat shock [13]. The cells were spread on agar plates containing ampicillin (100 μg/mL)/X-gal (40 μg/mL)/IPTG (0.5 mM) and incubated overnight at 37 °C for blue/white colonies. Clones were screened by streaking them on fresh ampicillin/X-gal/IPTG agar plates and incubated overnight at 37 °C. Plasmids were extracted from positive transformants using the Wizard® Plus SV Minipreps DNA purification System (Promega®) according to the manufacturer's instructions and restriction-digested with *Eco*RI to confirm that the insert DNA was present in the plasmid.

Table 1
Primers for partial cDNA amplification, RACE PCR and isolation of full length *CYP4H44P*.

Product	Primer	Sequence
CYP Fragments	Forward	5′-GATACGTTCATGTTTGAGGGGCA-3′
	Reverse	5′-GCGATCTTTTGGCCTATGC-3′
3′-RACE	3R1	5′-GTTCATGTTTGAGGGGCACGATA-3′
	3R1N	5′-AAGGAGGTTTTGCGGGTTTATC-3′
	Oligo(dT)$_{25}$V	5′-TTTTTTTTTTTTTTTTTTTTTTTTGCA-3′
5′-RACE	SGPR	5′-CTCCTTGATCACCATATCCAG-3′
CYP4H44P	Forward	5′-TAGCACCAGACCCGGGTCCAGCCAT-3′
	Reverse	5′-TTATCTGCTCTCGATTCTTAACGCCATATT-3′

2.6. Sequencing

Sanger sequencing was performed by First Base Laboratories Sdn Bhd with the universal primers SP6 and T7 using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems®) and separating DNA fragments in an ABI PRISM 3730xl Genetic Analyzer (Applied Biosystems®). Sequence similarity was performed in the National Centre for

Biotechnology Information (NCBI) database using Basic Local Alignment Search Tool (BLAST).

2.7. Verification of full length cDNA

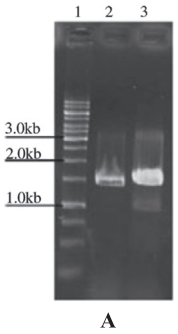
To verify that the 3′- and 5′-RACE products were from the same cDNA, the full length cDNA was amplified using the Advantage® GC 2 PCR Kit (Clontech®). The cycling condition was as follows; 94 °C/3 min, 30 cycles of 94 °C/20 s, 55 °C/20 s and 72 °C/2 min and finally 72 °C/10 min. Purification, cloning and sequencing of the product was as described in subsections 2.4 to 2.6.

2.8. Multiple sequence alignment and phylogenetic analysis

Sequence identity search was performed using BLAST. CYP sequences with homology to the query sequence were retrieved from the NCBI database and used for multiple sequence alignment with Clustal Omega [14]. The alignment file was used to compute pairwise evolutionary distances and construct a Maximum Likelihood phylogenetic tree based on the Tamura-Nei model [15] with 1000 bootstraps in MEGA6 [16].

2.9. Bioinformatics analysis

PSORT II (http://psort.hgc.jp/form2.html) was used to predict subcellular localisation of the deduced protein. Detection of potential signal peptide and transmembrane helix were also predicted using PrediSi (http://www.predisi.de/) and TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/). Phosphorylation and O-glycosylation sites were determined using NetPhos 2.0 Server (http://www.cbs.dtu.dk/services/NetPhos/) and DictyOGlyc 1.1 (http://www.cbs.dtu.dk/services/DictyOGlyc/) respectively, while GOR IV (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html) was used to predict the secondary structure. Homology modelling was performed with the automated modelling mode in SWISS-MODEL [17] to obtain the three dimensional (3D) model of the deduced protein. The model was subjected to structure validation using PROCHECK [18] and ProSA-web [19].



A

```
1     gtttccaaggggaaggggaggggcgaaaaattccccctctttccccctcggaaccgggt
63    gggcccatttgggcaactttgggaagtttcccaacaggagccttcgggaattttccccac
123   ggttttggatttgccaaacaaggggccaggatttggttacttaaggattgtttgggcga
183   tagcaccagacccgggtccagccatgggtccagatactcgtaaatgtagtttttcttggg
243   gaccttggacagcaaaacccgttcgatggccttccggctaaaaaatggaagctgaaactc
303   ttcggcgagaagtggttccaacggaggaagatccttaccccgacgttccattttaagatt
363   ttggagcagtttctggaggttttcaatgcggaaactgatcgattggtgacgaagatagag
423   cagcacgttggcggagaggagttcgatatgtatcagtatattacgttgcatgcactggac
                                      M  Y  Q  Y  I  T  L  H  A  L  D
483   agcatttgtgaaacttccatgggcgtatcgatcaatgcgctggacaacccggacaacgcg
       S  I  C  E  T  S  M  G  V  S  I  N  A  L  D  N  P  D  N  A
543   tacgttcacgccatcaaagactttggcagcatcgtcatccagcgaacgttctccgcgttg
       Y  V  H  A  I  K  D  F  G  S  I  V  I  Q  R  T  F  S  A  L
603   cgatcgtttccactactctacttcctgcatccattctactggcgccagcagaagctaatc
       R  S  F  P  L  L  Y  F  L  H  P  F  Y  W  R  Q  Q  K  L  I
663   aaaaccatgcacaatttcaccaactctgtaatcaaggccaaacggcaagccttggaggag
       K  T  M  H  N  F  T  N  S  V  I  K  A  K  R  Q  A  L  E  E
723   aaacgtcacaccgaaggtgaaaccaaggagcacaacgaagatgatggaatctacggtaag
       K  R  H  T  E  G  E  T  K  E  H  N  E  D  D  G  I  Y  G  K
783   aaacggatgagtttcctggacctgctgctgaacgagagctccatgagcgatgcggacatc
       K  R  M  S  F  L  D  L  L  L  N  E  S  S  M  S  D  A  D  I
843   cgcgaggaggtcgatacgttcatgttcgaaggacacgatacgaccacttcggggatttac
       R  E  E  V  D  T  F  M  F  E  G  H  D  T  T  T  S  G  I  Y
903   ttctcgttgatggcccttaccatgcatccggatattcaggaaaggttgtacggtgagatt
       F  S  L  M  A  L  T  M  H  P  D  I  Q  E  R  L  Y  G  E  I
963   cggcaggtgttggaaacgggagaggagcgacatgctccgctaaccaacgcaacgctgcaa
       R  Q  V  L  E  T  G  E  E  R  H  A  P  L  T  N  A  T  L  Q
1023  cagatgaagtacctggatatggtgatcaaggaggttttgcgggtttatccatcggtgccg
       Q  M  K  Y  L  D  M  V  I  K  E  V  L  R  V  Y  P  S  V  P
1083  atcatcggccgggagctgctggaggacgttgagatcaacggatgccaagtaccgcgggga
       I  I  G  R  E  L  L  E  D  V  E  I  N  G  C  Q  V  P  R  G
1143  accgccatggtggtgatcatccataatgtgcaccgcaacgcggaagtgttccccgatccg
       T  A  M  V  V  I  I  H  N  V  H  R  N  A  E  V  F  P  D  P
1203  gagcggttcgatccggaaagattttccgacgaatccggcgggaaaagagggccgtacgat
       E  R  F  D  P  E  R  F  S  D  E  S  G  G  K  R  G  P  Y  D
1263  tacataccgttcagtgtgggagcgcgaaattgcatcggtcaaaagtacgctttactggag
       Y  I  P  F  S  V  G  A  R  N  C  I  G  Q  K  Y  A  L  L  E
1323  atgaaggttacgctggtgaaactgctgttggcatataggttcattccggggaagtctacg
       M  K  V  T  L  V  K  L  L  L  A  Y  R  F  I  P  G  K  S  T
1383  gattcgattcgtattcagggagaccttgtgttgcgtccgttcgggaatatggcgttaaga
       D  S  I  R  I  Q  G  D  L  V  L  R  P  F  G  N  M  A  L  R
1443  atcgagagcagataaagttgtgggtgtgcaaggattgaaatgcaaataaaaagctgttcg
       I  E  S  R
1503  cttgggtgcaaaaaaaaaaaaaaaaaaaaaaaaa
```

B

Figure 1. Gel electrophoresis of PCR products and sequence of *CYP4H44P*.
(A) Amplified products after PCR to verify full length cDNA. Lane 1: 1 kb DNA ladder. Lane 2: *CYP4H44P* cDNA. Lane 3: CYP fragments. (B) *CYP4H44P* sequence consisting of 1 537 nucleotides coding for 335 amino acids. Motifs found in CYP and CYP family 4 genes have been underlined.

3. Results

3.1. Sequence and homology of pseudogene

Amplification of the full length cDNA yielded a product of approximately 1.5 kb (Figure 1A). Sequencing of the product after cloning and plasmid extraction showed that it has 1 537 nucleotides (nt) with a deduced open reading frame of 1 008 nt, which translated into 335 amino acids (aa) (Figure 1B). The sequence was classified as a pseudogene and assigned the name *CYP4H44P* by the P450 Nomenclature Committee. It has 449 and 80 nt in the 5′-and 3′-Untranslated Regions, a stop codon (TAA) starting at the 1 455th nt position, a polyadenylation signal (AATAAA) and a 26 nt long poly A tail. The translated sequence contains CYP motifs like ExxR ($E^{202}VLR^{205}$), PERF ($P^{256}ERF^{259}$), FxxGxxxCxG ($F^{275}SVGARNCIG^{284}$) and the 13-residue sequence ($E^{134}VDTFMFEGHDTT^{146}$) of family 4 CYPs (Figure 1B).

The nt sequence of *CYP4H44P* is 99% identical with *CYP4H28* [Query cover (QC) = 93%] and *CYP4H28v2* (QC = 99%). The *CYP4H44P* aa sequence also has an identity of 99% with *CYP4H28* (QC = 100%) and *CYP4H28v2* (QC = 100%).

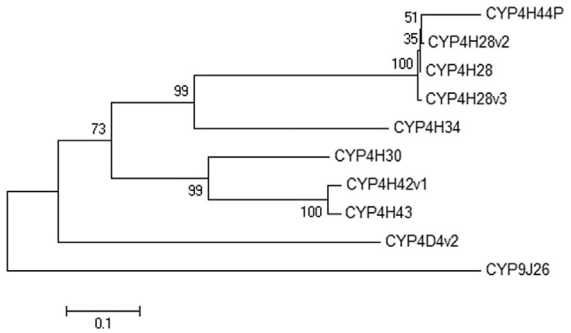


Figure 2. Molecular phylogenetic analysis conducted in MEGA6 [17] using the Maximum Likelihood method based on the Tamura-Nei model [16].

The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 10 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Non-coding. All positions containing gaps and missing data were eliminated. There were a total of 1380 positions in the final dataset. Species name and GenBank accession number of the CYPs are as follows; *CYP4H28* (*Ae. aegypti*, XM_001656716), *CYP4H28v2* (*Ae. aegypti*, KC481237), *CYP4H28v3* (*Ae. aegypti*, KF779931), *CYP4H30* (*Ae. aegypti*, XM_001656715), *CYP4H34* (*Cx. quinquefasciatus*, JQ001927), *CYP4H42v1* (*Ae. albopictus*, KF029763), *CYP4H43* (*Ae. albopictus*, KF029765), *CYP4D4v2* (*Musca domestica*, EF615001), *CYP9J26* (*Ae. aegypti*, XM_001649047).

3.2. Evolutionary relationship

The phylogenetic tree (Figure 2) showed the relationship between the pseudogene and other family 4 CYP genes with *CYP9J26* as outgroup. The closest evolutionary relationship of the pseudogene was with the *CYP4H28* alleles, clustering with a bootstrapping support of 100%. *CYP4H44P* and the *CYP4H28* group were closely related to *CYP4H34* of *Culex quinquefasciatus* (*Cx. quinquefasciatus*) (99% bootstrap value) than the cluster of *CYP4H42v1* and *CYP4H43* of *Aedes albopictus* (*Ae. albopictus*) and *CYP4H30* of *Cx. quinquefasciatus* (73% bootstrap value). The computed estimated evolutionary divergence of *CYP4H44P* from the other sequences indicated that it was least diverged from *CYP4H28* and *CYP4H28v2* (Table 2).

3.3. Bioinformatics analysis

PSORT II software predicted that the CYP4H44P-deduced protein had the following probabilities of localisation; 60.9% (cytoplasmic), 17.4% (nuclear), 13.0% (mitochondrial), 4.3% (vacuolar) and 4.3% (vesicles of secretory system). Signal peptide, transmembrane helix and O-glycosylated sites were not predicted in the deduced protein (Figure 3). However, sixteen (16) phosphorylated sites were predicted in the deduced protein (Figure 3). The predicted secondary structure was made up of alpha helices (35.52%), random coils (47.46%) and extended strands (17.01%). The 3D structure was modelled using CYP3A4 (PDB ID: 4D6Z) as template. This template was a human CYP and had 34.74% sequence identity with CYP4H44P. Ramachandran plot analysis of the model showed that 85.2, 12.0, 2.1 and 0.7% of the residues were in the most favoured, additional allowed, generously allowed and disallowed regions respectively (Figure 4). Validation of the model using the ProSA z-score indicated a model of good quality with a value of −5.81 (Figure 4).

Table 2
Estimated evolutionary divergence between CYP sequences.

CYP genes	1	2	3	4	5	6	7	8	9
1.*CYP4H44P*									
2.*CYP4H28v2*	0.084								
3.*CYP4H28*	0.084	0.006							
4.*CYP4H28v3*	0.088	0.011	0.007						
5.*CYP4H30*	0.717	0.655	0.650	0.653					
6.*CYP4H34*	0.630	0.568	0.562	0.560	0.618				
7.*CYP4H42v1*	0.748	0.693	0.689	0.696	0.348	0.625			
8.*CYP4H43*	0.759	0.699	0.695	0.705	0.341	0.630	0.036		
9.*CYP4D4v2*	0.880	0.817	0.810	0.825	0.771	0.837	0.722	0.752	
10.*CYP9J26*	1.191	1.146	1.137	1.145	1.017	1.104	1.058	1.057	1.127

The number of base substitutions per site from between sequences are shown. Analyses were conducted using the Tamura-Nei model [16]. The analysis involved 10 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 1 380 positions in the final dataset.

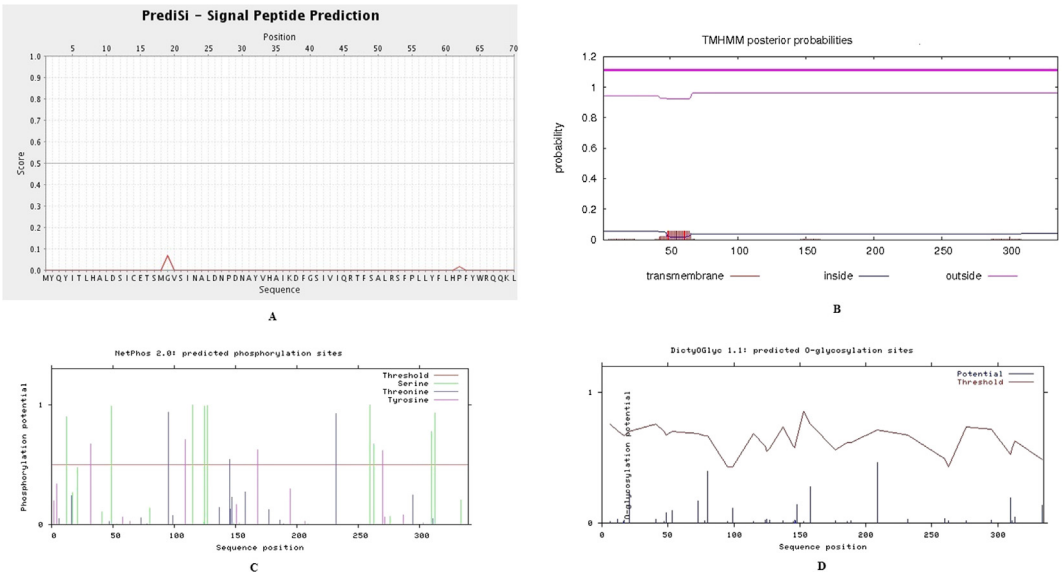


Figure 3. Predicted features in CYP4H44P-deduced protein.
(A) Signal peptide; (B) Transmembrane helix; (C) Phosphorylated sites; (D) O-glycosylated sites.

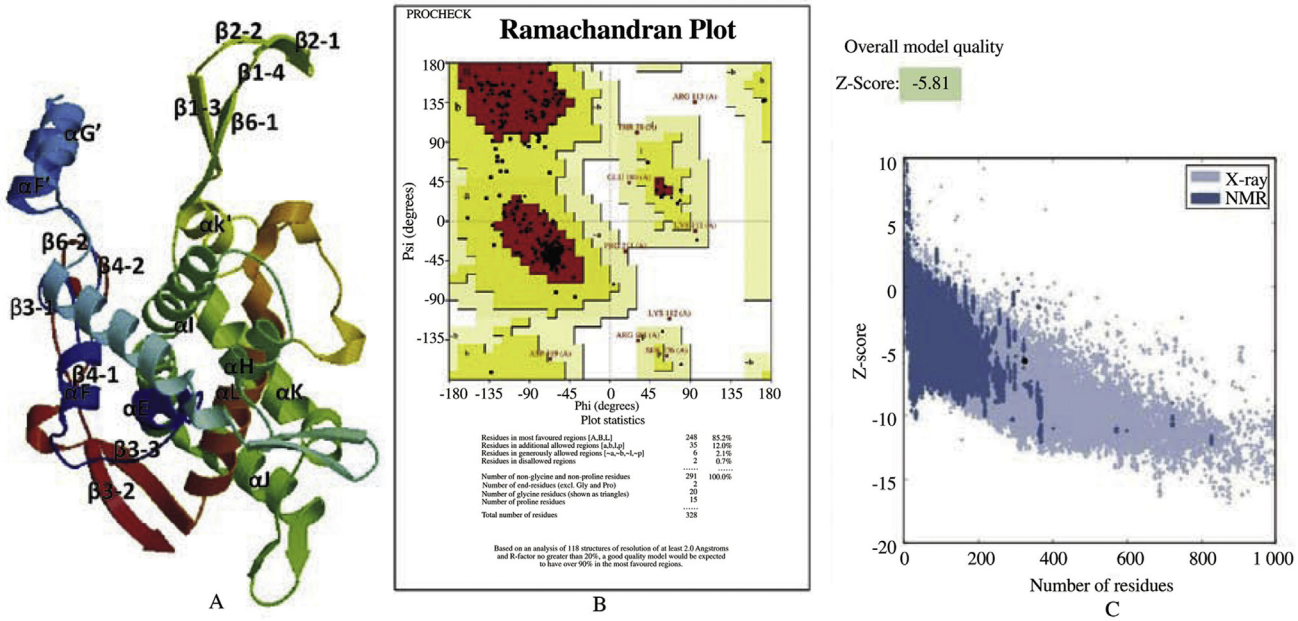


Figure 4. 3D model and validation results of CYP4H44P-deduced protein.
(A) 3D model with labelled secondary structures; (B) Ramachandran plot analysis; (C) ProSA z-score.

4. Discussion

This study isolated a processed CYP pseudogene, *CYP4H44P*, from *Ae. aegypti* larvae. The *CYP4H44P* sequence has been deposited at the GenBank database with the accession number KF779932. The translated sequence has several CYP motifs including the signature haem-binding motif FxxGxxxCxG (F^{275}SVGARNCIG^{284}) [1,20] but lacks the WxxxR motif due to its truncated N-terminal region. The WxxxR motif is presumed to interact with the propionate of haem to form a charge pair [1]. However, translation of the 5′-Untranslated Region shows that nt sequences from the 315–329th position code for this motif but this is upstream of the start codon (A^{450}TG^{452}) in this sequence. *CYP4H44P* is highly identical to *CYP4H28* and *CYP4H28v2* but has a shorter aa sequence in the coding region than the other functional products due to its truncated 5′ region. Truncation and structural loss in *CYP4H44P* might have led to its classification as a pseudogene.

Phylogenetic and evolutionary divergence analyses indicated that *CYP4H44P* is closely related to the *CYP4H28* group. Genes such as *CYP4H28* and *CYP4H28v2* are induced by xenobiotics [10,21] so the close evolutionary relationship between these and *CYP4H44P* suggests that *CYP4H44P* could have possessed a similar trait if it was a functional gene. The high sequence identity and evolutionary relatedness of *CYP4H44P* to the functional *CYP4H28* alleles imply that divergence of the pseudogene is likely to be recent [22]. Five CYP pseudogenes have been identified in the *Ae. aegypti* genome but none belongs to the family 4 CYP group [3]. To the authors' best

knowledge, this is the first family 4 CYP pseudogene from *Ae. aegypti*. The family 4 CYP is a functionally diverse group with a broad range of functions [1] and although *CYP4H44P* may presumptively be enzymatically non-functional, it will be interesting to determine its biological significance since biological functions have been reported for some pseudogenes [7].

The predicted localisation of the pseudogene-deduced protein to the cytoplasm is similar to the predicted cytosolic localisation of pseudogene-deduced proteins in Shidhi *et al* [9]. The protein was also predicted to have several phosphorylated sites. Phosphorylation of functional CYPs regulates protein activity [23]; hence, phosphorylation may help to modulate activity of the pseudogene-deduced protein if it was synthesised. The fold of the 3D model is akin to the conserved fold of CYPs [20]. The model has 0.7% of its residues in the disallowed regions, which is comparable to predicted CYP models of *Tribolium castaneum*, which had between 1.1% and 2.2% of residues in the disallowed regions [24]. Quality assessment using validation tools indicated that the model was of good quality and compares well with predicted models of proteins deduced from pseudogenes [9].With the predicted possibility of synthesising functional proteins from pseudogenes [9], the predicted characterisation of the pseudogene in this study offers useful insight in this regard.

In conclusion, a processed CYP pseudogene (*CYP4H44P*) was isolated from *Ae. aegypti*. The pseudogene is evolutionary related with *CYP4H28* alleles and appears to have recently diverged from this group. Bioinformatics tools were used to characterise the deduced protein and the predicted 3D model indicates that it has the conformational fold of CYPs. Since pseudogenes may not be functionally defunct as previously thought, identification of this pseudogene provides a platform for investigating its functional role in the mosquito.

Conflict of interest statement

We declare that we have no conflict of interest.

Acknowledgements

This work was funded by the Malaysian Ministry of Education (ERGS 203/PJJAUH/6730097) and Universiti Sains Malaysia (RU grant 1001/PJJAUH/815095).

References

[1] Feyereisen R. Insect CYP genes and P450 enzymes. In: Gilbert LI, editor. *Insect molecular biology and biochemistry*. Oxford: Elsevier; 2012, p. 236-316.
[2] Nelson DR. A world of cytochrome P450s. *Philos Trans R Soc Lond B Biol Sci* 2013; **368**(1612): 20120430; http://dx.doi.org/10.1098/rstb.2012.0430.
[3] Issa MS. *Molecular characterization and functional analysis of cytochrome P450 genes in the yellow fever mosquito Aedes aegypti (Diptera: Culicidae)*. Manhattan: Kansas State University; 2014. MSc thesis.
[4] Mighell AJ, Smith NR, Robinson PA, Markham AF. Vertebrate pseudogenes. *FEBS Lett* 2000; **468**(2–3): 109-114.
[5] Balakirev ES, Ayala FJ. Pseudogenes: are they "junk" or functional DNA? *Annu Rev Genet* 2003; **37**: 123-151; http://dx.doi.org/10.1146/annurev.genet.37.040103.103949.
[6] Tutar Y. Pseudogenes. *Comp Funct Genomics* 2012; **2012**: 424526; http://dx.doi.org/10.1155/2012/424526.
[7] Li W, Yang W, Wang XJ. Pseudogenes: pseudo or real functional elements? *J Genet Genomics* 2013; **40**(4): 171-177.
[8] Guo X, Lin M, Rockowitz S, Lachman HM, Zheng D. Characterization of human pseudogene-derived non-coding RNAs for functional potential. *PLoS One* 2014; **9**(4): e93972; http://dx.doi.org/10.1371/journal.pone.0093972.
[9] Shidhi PR, Suravajhala P, Nayeema A, Nair AS, Singh S, Dhar PK. Making novel proteins from pseudogenes. *Bioinformatics* 2015; **31**(1): 33-39.
[10] Saavedra-Rodriguez K, Strode C, Flores AE, Garcia-Luna S, Reyes-Solis G, Ranson H, et al. Differential transcription profiles in *Aedes aegypti* detoxification genes after temephos selection. *Insect Mol Biol* 2014; **23**(2): 199-215.
[11] Elgarj FNA, Wajidi MFF. Molecular cloning and characterization a novel gene encoding CYP4H28v2 from the mosquito, *Aedes aegypti*. *Int J Chem Environ Biol Sci* 2013; **1**(2): 240-243.
[12] El-garj FMA, Avicor SW, Wajidi MFF, Jaal Z. Comparative efficacy of spatial repellents containing d-allethrin and d-trans allethrin against the major dengue vector *Aedes aegypti* (Linnaeus). *Asian Biomed* 2015; **9**(3): 313-320.
[13] Chung CT, Niemela SL, Miller RH. One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci USA* 1989; **86**(7): 2172-2175.
[14] McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, et al. Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res* 2013; **41**(Web Server issue): W597-W600; http://dx.doi.org/10.1093/nar/gkt376.
[15] Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 1993; **10**(3): 512-526.
[16] Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 2013; **30**(12): 2725-2729.
[17] Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucl Acids Res* 2014; **42**(Web Server issue): W252-W258; http://dx.doi.org/10.1093/nar/gku340.
[18] Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK – a program to check the stereochemical quality of protein structures. *J App Cryst* 1993; **26**(2): 283-291.
[19] Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucl Acids Res* 2007; **35**(Web Server issue): W407-W410; http://dx.doi.org/10.1093/nar/gkm290.
[20] Werck-Reichhart D, Feyereisen R. Cytochromes P450: a success story. *Genome Biol* 2000; **1**(6); http://dx.doi.org/10.1186/gb-2000-1-6-reviews3003. reviews3003.1–reviews3003.9.
[21] El-garj FMA, Avicor SW, Wajidi MFF. Xenobiotic-induced expression of detoxification genes, *CYP4H28v2* and *CYP4H31v2* in the dengue mosquito *Aedes aegypti*. *Trop Biomed* 2016. in press.
[22] Wilde CD. Pseudogenes. *CRC Crit Rev Biochem* 1986; **19**(4): 323-352.
[23] Lamb DC, Waterman MR. Unusual properties of the cytochrome P450 superfamily. *Philos Trans R Soc Lond B Biol Sci* 2013; **368**(1612): 20120434; http://dx.doi.org/10.1098/rstb.2012.0434.
[24] Zhu F, Moural TW, Shah K, Palli SR. Integrated analysis of cytochrome P450 gene superfamily in the red flour beetle, *Tribolium castaneum*. *BMC Genomics* 2013; **14**: 174; http://dx.doi.org/10.1186/1471-2164-14-174.